CHAPTER 31

Pairs of Cameras

31.1 GEOMETRY

Place two perspective cameras in 3D, and construct the line segment joining their focal points. This line segment is known as the *baseline*. Now extend the baseline to a line. This line intersects each image plane in an important point, known as the *epipole* for that image plane in that configuration of cameras (Figure 31.1).

The epipole in a single image reveals a great deal of information about the camera has moved. For intuition, construct some of the lines through the epipole, then compare to Figure 31.1. Figure 31.2 shows some examples.

31.1.1 Epipoles, Epipolar Lines and Correspondence

If the cameras view some point $\mathbf{X}$ into 3D, camera 1 sees that point at $\mathbf{x}_1$ and camera 2 sees that point at $\mathbf{x}_2$. The points $\mathbf{X}$, $\mathbf{f}_1$ and $\mathbf{f}_2$ define a plane in 3D. This plane intersects the image plane of camera 1 in a line I shall call $\mathbf{l}_1(\mathbf{X})$. Similarly, it intersects camera 2 in an line $\mathbf{l}_2(\mathbf{X})$. These lines must pass through their respective epipoles (Figure **??**), and are known as *epipolar lines*.

If you see a point in the first camera at $\mathbf{x}_1$, you will need to find $\mathbf{x}_2$ to produce a 3D reconstruction. But not any point in camera 2 could correspond to $\mathbf{x}_1$. If you know enough about relative camera configuration, you can construct the epipolar line $\mathbf{l}_2$ in the second image, and $\mathbf{x}_2$ must lie on this line.

Introduce a second point $\mathbf{Y}$ that doesn't lie on the plane through $\mathbf{X}$, $\mathbf{f}_1$ and $\mathbf{f}_2$. This yields a second plane, which produces its own epipolar lines (Figure 31.1). The baseline defines a whole family of planes that contain the baseline; this is often referred to as a *star* of planes).

31.1.2 The Fundamental Matrix

As Figure 31.1 shows, a point $\mathbf{X}$ in 3D selects a plane from the family of planes through both focal points, and this plane intersects each image plane in epipolar lines. You can identify this plane using $\mathbf{x}_1$, so the figure illustrates a mapping from points in camera 1 to lines through the epipole in camera 2. The mapping works like this: Select a point in camera 1; construct the plane through this point and the two focal points; now intersect that plane with camera 2's image plane; and you have the corresponding epipolar line. The projection of $\mathbf{X}$ into camera 2 ($\mathbf{x}_2$) *must* lie on this line.

This construction exposes an extremely important relationship between $\mathbf{x}_1$ and $\mathbf{x}_2$. Each of these points lies on its image plane, and so has three homogeneous coordinates. However, you can express these points in 3D. Write $\mathbf{X}_1$ for the point $\mathbf{x}_1$ written in four homogeneous coordinates, and notice that there is some 4×3

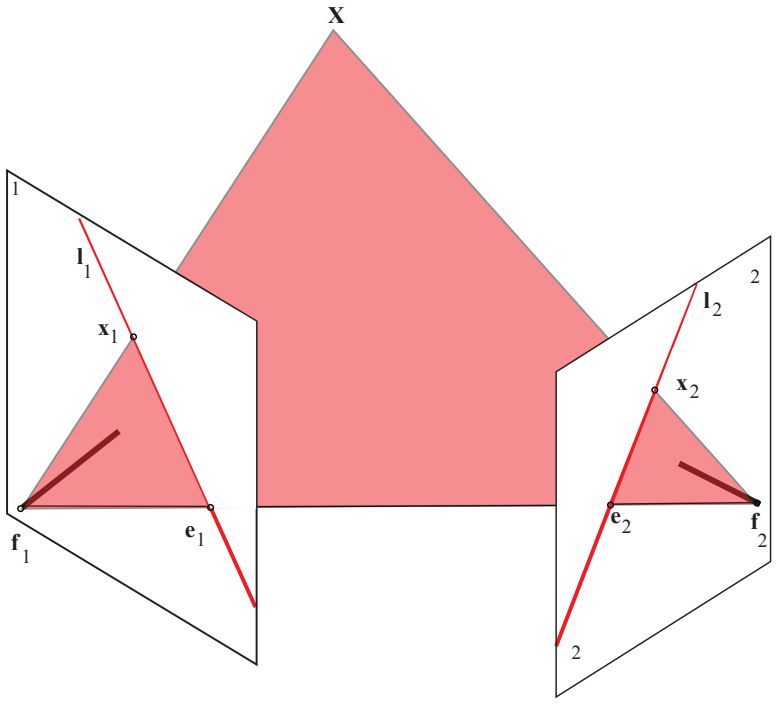


FIGURE 31.1: *Two cameras view a point* $\mathbf{X}$ *in space. The epipoles are formed by the line connecting the focal points (*$\mathbf{f}_1$ *and* $\mathbf{f}_2$*) of the two perspective cameras (the baseline). This line intersects camera* 1*'s image plane in* $\mathbf{e}_1$ *(the epipole in the first image), and camera* 2*'s image plane in* $\mathbf{e}_2$ *(the epipole in the second image). As long as the two focal points are distinct, the epipoles are properly defined, although they may appear far outside the image (or even at infinity). The point* $\mathbf{X}$ *projects to* $\mathbf{x}_1$ *in camera* 1 *and* $\mathbf{x}_2$ *in camera* 2*. The three points* $\mathbf{f}_1$*,* $\mathbf{f}_2$ *and* $\mathbf{X}$ *define a plane in 3D. This plane contains the baseline, and intersects camera* 1*'s image plane in a line (*$\mathbf{l}_1$ *in the figure) that passes through* $\mathbf{e}_1$ *and camera* 2*'s image plane in a line (*$\mathbf{l}_2$ *in the figure) that passes through* $\mathbf{e}_2$*.*

matrix $\mathcal{P}_1$ so that $\mathbf{X}_1 = \mathcal{P}_1\mathbf{x}_1$ (**exercises**).

The four points $\mathbf{X}_1$, $\mathbf{f}_1$, $\mathbf{f}_2$ and $\mathbf{X}_2$ lie on the same plane. Each is a point in 3D. If you write each point in homogenous coordinates, you must have

$$\text{determinant}([\mathbf{X}_1, \mathbf{f}_1, \mathbf{f}_2, \mathbf{X}_2]) = 0.$$

(**exercises**). In turn, this means that there is some 4×4 matrix $\mathcal{G}$ such that

$$\mathbf{X}_1^T \mathcal{G} \mathbf{X}_2 = 0.$$

But $\mathbf{X}_1 = \mathcal{P}_1\mathbf{x}_1$, etc. So there is some 3×3 matrix $\mathcal{F}$ such that

$$\mathbf{x}_1^T \mathcal{F} \mathbf{x}_2 = 0.$$

This matrix is known as the *fundamental matrix*.

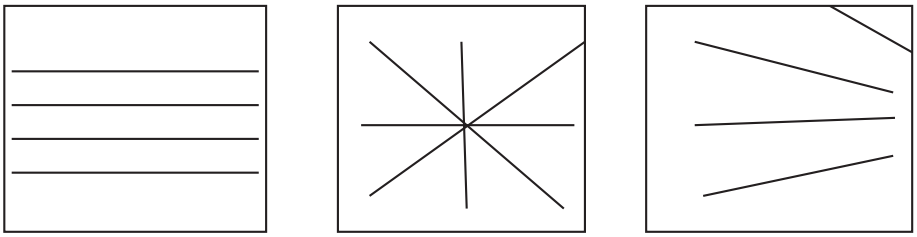


FIGURE 31.2: *The epipoles reveal information about how cameras have moved. This is most easily seen by thinking about the epipolar lines. These three examples sketch the underlying intuition, and should be looked at together with Figure 31.1. On the* **left**, *the camera has translated parallel to the image plane;* **center**, *the camera has moved perpendicular to the image plane; and* **right**, *the camera has translated parallel to the image plane then rotated slightly. You should check each case carefully.*

There isn't any particular reason the cameras are labelled 1 and 2 – you could swap the labels, without affecting the geometry. Notice that $\mathbf{x}_1^T \mathcal{F} \mathbf{x}_2 = 0$ implies that ${\mathbf{x}_2}^T {\mathcal{F}}^T \mathbf{x}_1 = 0$. This means that, if you swap the labels, you transpose the fundamental matrix. In turn, a procedure that finds something in camera 2 using data from camera 1 can also be used to find something in camera 1 using data from camera 2.

Remember this: *For any pair of cameras which do not share a focal point, there is a fundamental matrix $\mathcal{F}$ with the property that for any pair $\mathbf{x}_1$, $\mathbf{x}_2$, where $\mathbf{x}_1$ is the image of a 3D point in the first camera and $\mathbf{x}_2$ is the image of that point in the second camera,*

$$\mathbf{x}_1^T \mathcal{F} \mathbf{x}_2 = 0$$

31.1.3 Properties of the Fundamental Matrix

The fundamental matrix for a pair of cameras reveals epipolar lines. Choose some point $\mathbf{x}_1$ in the first camera. Now for *every* point $\mathbf{x}'$ in camera 2 that could match $\mathbf{x}_1$,

$$\mathbf{x}_1^T \mathcal{F} \mathbf{x}' = \left(\mathbf{x}_1^T \mathcal{F}\right) \mathbf{x}' = 0$$

and you can think of ${\mathcal{F}}^T \mathbf{x}_1$ as a vector containing the coefficients of a line. This line is the epipolar line corresponding to $\mathbf{x}_1$. You can think of the fundamental matrix as a map from points in one camera to lines in the other camera.

Procedure: 31.1 *Obtaining an epipolar line from a fundamental matrix*

The epipolar line in camera 2 corresponding to $\mathbf{x}_1$ in camera 1 consists of the set of points $\mathbf{x}'$ in camera 2 which satisfy the equation

$$\left(\mathbf{x}_1^T \mathcal{F}\right) \mathbf{x}' = 0$$

and the coefficients of the line are

$$\left(\mathcal{F}^T \mathbf{x}_1\right).$$

It follows that the coefficients of the epipolar line in camera 1 corresponding to $\mathbf{x}_2'$ in camera 2 are

$$\left(\mathcal{F} \mathbf{x}_2'\right).$$

The fundamental matrix for a pair of cameras reveals epipoles. Every epipolar line passes through the epipole, so the epipole must be the point in camera 2, $\mathbf{e}_2$, such that for *any* choice of $\mathbf{x}_1$, $\left(\mathbf{x}_1^T \mathcal{F}\right) \mathbf{e}_2 = 0$. The only way to achieve this is if

$$\mathcal{F} \mathbf{e}_2 = \mathbf{0}$$

so the epipole in camera 2 is the right null vector of $\mathcal{F}$. Similarly, the epipole in camera 1 is the left null vector of $\mathcal{F}$. This means that $\mathcal{F}$ cannot have full rank.

Procedure: 31.2 *Obtaining epipoles from a fundamental matrix*

The epipole in camera 2 is the point $\mathbf{e}_2$ such that

$$\mathcal{F} \mathbf{e}_2 = \mathbf{0}.$$

It follows that the epipole in camera 1 is the point $\mathbf{e}_1$ such that

$$\mathcal{F}^T \mathbf{e}_1 = \mathbf{0}.$$

In fact, $\mathcal{F}$ must have rank 2. The rank can't be three, because there are epipoles. The fundamental matrix is a map from points in one camera to lines in the other. (**exercises**). If the rank were 0, the fundamental matrix would map any point in one camera to a zero vector (this happens if the camera is not translated, a rather special case, Section **??**). Rank 2 is the only available alternative.

Remember this: *The fundamental matrix $\mathcal{F}$ must have rank 2 unless the two cameras share a focal point, when it consists of zeros.*

Since

$$\mathbf{x}_1^T \mathcal{F} \mathbf{x}_2 = 0 = \mathbf{x}_1^T (s\mathcal{F}) \mathbf{x}_2$$

for any $s \neq 0$, the fundamental matrix is only really meaningful up to scale.

> **Remember this:** *The fundamental matrix $\mathcal{F}$ is only meaningful up to scale. You should think of this matrix as a point in 8 dimensional space represented with 9 homogenous coordinates.*

31.1.4 Estimating the Fundamental Matrix

The fundamental matrix can be estimated from point correspondences. Assume two cameras view a set of points $\mathbf{X}_i$ in 3D. Write $\mathbf{x}_{1;i}$ for the image of the i'th point in camera 1 and $\mathbf{x}_{2;i}$ for the image of the i'th point in camera 2. Each pair yields one equation that constrains the fundamental matrix, that is

$$\mathbf{x}_{1;i}^T \mathcal{F} \mathbf{x}_{2;i} = 0$$

(remember – you *know* the coordinates of the point in each camera, so the unknowns here are the elements of $\mathcal{F}$). The equation can be written out in detail if you have sufficient tolerance for notation. Recall $\mathbf{x}_{1;i}$ is in homogeneous coordinates, so it has three coefficients (write $x_{u,1;i}$ for the u'th coefficient of $\mathbf{x}_{1;i}$; the third coefficient might well be 1). Then you have

$$\sum_{uv} (x_{u,1;i} x_{v,2;i}) f_{uv} = 0.$$

There is one equation for each pair of points, and each equation is homogenous, so eight pairs of points yield an estimate of $\mathcal{F}$. If you happen to have more points, solve with least squares. The estimate is up to scale, but the fundamental matrix is only meaningful up to scale. The procedure is known as the *8 point algorithm*.

There are useful improvements available. The scale of the image coordinate system can have a real effect on the estimate of $\mathcal{F}$, because squared terms appear in the matrix. A more accurate estimate of $\mathcal{F}$ can be obtained by scaling the image so that the largest coordinate direction runs from 0 to 1 (rather than, say, 0 to 768).

Further, the eight point algorithm does not produce a rank 2 estimate of the fundamental matrix has rank 2. Improve the estimate $\mathcal{F}^*$ by finding the closest rank 2 matrix. Do this using a singular value decomposition, as below.

Procedure: 31.3 *Estimating the fundamental matrix*

Scale the image coordinate system for camera 1 and camera 2 so that the largest coordinate direction runs from 0 to 1. Find 8 pairs of corresponding points $\mathbf{x}_{1,i}$ and $\mathbf{x}_{2,i}$. Now solve the system of 8 homogenous equations given by

$$\mathbf{x}_{1,i}^T \mathcal{F}^* \mathbf{x}'_{2,i} = 0$$

in $\mathcal{F}^*$. Decompose $\mathcal{F}^*$ as $\mathcal{U}\Sigma\mathcal{V}^T$. Set the smallest singular value in Σ to 0, to obtain $\Sigma_{\text{rank 2}}$. The estimate is then

$$\hat{\mathcal{F}} = \mathcal{U}\Sigma_{\text{rank 2}}\mathcal{V}^T.$$

The estimation procedure requires 8 corresponding pairs. The natural source of these pairs is RANSAC, but notice that this is a large number of pairs, so you really do not want to just select from all pairs of points. Instead, find and describe interest points using (for example) the methods of Chapter **??**, and use only pairs whose descriptors match well.

The rank 2 constraint on $\mathcal{F}$ can be exploited further. The constraint is cubic in the coefficients of $\mathcal{F}$ (the constraint is $\det(\mathcal{F}) = 0$). Exploiting the constraint makes it possible to estimate a fundamental matrix with seven corresponding pairs, *if* you are willing to form the roots of a cubic. This isn't for everyone; details in [].

31.2 COORDINATE GEOMETRY

The drawings of the previous section illustrate geometric facts that do not depend on coordinates, but usually you need to use two images to reconstruct a point in 3D. This problem has to be worked in coordinates. Assume each camera has known intrinsics. This means you can calibrate each camera so that the coordinates the camera reports are the coordinates of a point in the camera's coordinate system and each has focal length 1.

31.2.1 Triangulating a Point in 3D from Two Images

Now choose a coordinate system so that the first camera has focal point at the origin, looks down the z-axis, and has image plane at $z = 1$, as in Section 29.2. To get the second camera, rotate the first camera by $\mathcal{R}^T$, then translate it by $\mathbf{t}$, so that $\mathbf{f}_2 = \mathbf{t}$. Notice that this means that a point at $\mathbf{X} = [X_1, X_2, X_3]^T$ in the first camera's coordinate system appears at $\mathbf{X}' = \mathcal{R}(\mathbf{X} - \mathbf{t})$ (if the camera rotates left, then all the points in the image frame move right). Figure 31.3 shows this setup.

I will write points in the second camera's coordinate system with a prime, and I will work this problem in affine coordinates. Write

$$\mathcal{R} = \begin{bmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \\ \mathbf{r}_3^T \end{bmatrix}$$

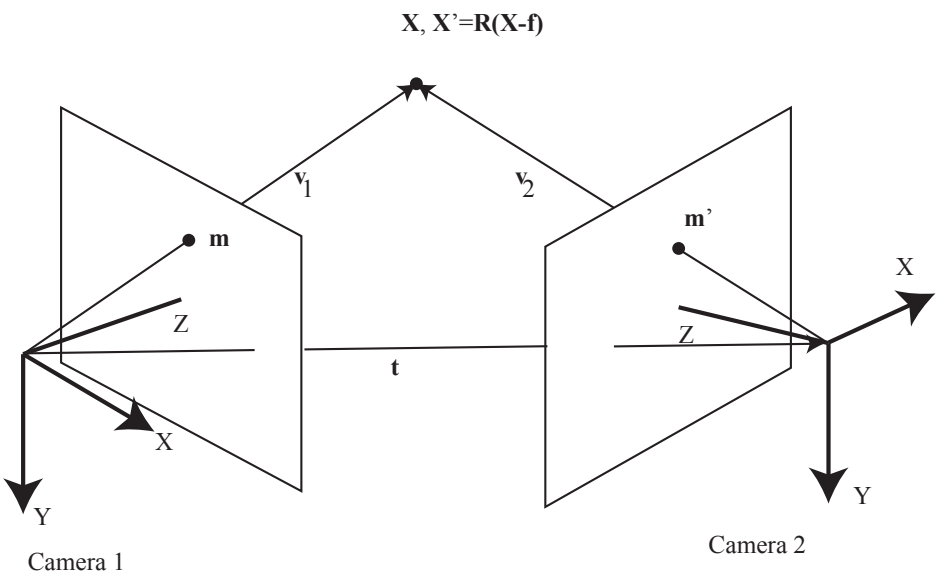


FIGURE 31.3: *Camera 1 is in the canonical configuration of Section 29.2 (focal point at the origin, looking down z-axis, image plane at z=1). Camera 2 is obtained from camera 1 by a rotation and translation. Each views a point, which is* $\mathbf{X}$ *in camera 1's coordinate system,* $\mathbf{X}' = \mathcal{R}(\mathbf{X} - \mathbf{f})$ *in camera 2's coordinate system. This point projects to* $\mathbf{m}$ *in camera 1, measured in its coordinate system and* $\mathbf{m}'$ *in camera 2, measured in its coordinate system. From* $\mathbf{m}$, $\mathbf{m}'$, $\mathcal{R}$ *and* $\mathbf{f}$ *you can recover* $\mathbf{X}$ *(or* $\mathbf{X}'$*) using the procedures in the text.*

You see the point in the first camera at

$$\mathbf{m} = \begin{bmatrix} m_1 \\ m_2 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{X_1}{X_3} \\ \frac{X_2}{X_3} \\ 1 \end{bmatrix}$$

and in the second camera at

$$\mathbf{m}' = \begin{bmatrix} m_1' \\ m_2' \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{\mathbf{r}_1^T(\mathbf{X}-\mathbf{t})}{\mathbf{r}_3^T(\mathbf{X}-\mathbf{t})} \\ \frac{\mathbf{r}_2^T(\mathbf{X}-\mathbf{t})}{\mathbf{r}_3^T(\mathbf{X}-\mathbf{t})} \\ 1 \end{bmatrix}.$$

For the moment, assume that $\mathbf{m}$, $\mathbf{m}'$, the rotation and the translation are all known exactly. Then

$$X_3\mathbf{m} = \mathbf{X}$$

and the only unknown is X_3. You can write two linear equations in this unknown,

which are

$$\begin{aligned} m_1'(\mathbf{r}_3^T(X_3\mathbf{m}-\mathbf{t})) - (\mathbf{r}_1^T(X_3\mathbf{m}-\mathbf{t})) &= 0 \\ m_2'(\mathbf{r}_3^T(X_3\mathbf{m}-\mathbf{t})) - (\mathbf{r}_2^T(X_3\mathbf{m}-\mathbf{t})) &= 0. \end{aligned}$$

These equations have to be consistent, which means that there is a relationship between $\mathbf{m}$ and $\mathbf{m}'$ that depends on $\mathcal{R}$ and $\mathbf{t}$. The relationship expresses the mapping from points to lines of the previous section. It could be obtained by some aggressive linear algebra, but is better constructed directly, which I do in the next section.

A warning: it is not a good idea to *estimate* X_3 using the equations above, because you will never actually know $\mathbf{m}$ and $\mathbf{m}'$ exactly. They are useful only to establish that you can recover X_3.

31.2.2 Triangulation by Minimization

Now assume you know $\mathcal{R}$ and $\mathbf{t}$, the intrinsic calibration matrices $\mathcal{K}$ and $\mathcal{K}'$, and have estimated locations $\mathbf{m}$ and $\mathbf{m}'$ for a pair of points that correspond. These estimates may not be exact – for example, they might come from an interest point matcher – but any error is small. You must recover the point in 3D.

The first camera has camera matrix $\mathcal{K}\mathcal{C}_p$. The second camera has camera matrix

$$\mathcal{K}'\left[\mathcal{R}|-\mathcal{R}\mathbf{t}\right]$$

(recall notation from Section 36.3, and check that this camera has focal point at $\mathbf{t}$). Now write $\mathbf{X} = [X_1, X_2, X_3]$ for a point in 3D in affine coordinates. The residual vector in camera 1 is the vector from the projection of $\mathbf{X}$ to $\mathbf{m}$, so

$$\mathbf{e}_1(\mathbf{X}) = \begin{bmatrix} k_{11}\frac{X_1}{X_3} + k_{12}\frac{X_2}{X_3} + k_{13}\frac{1}{X_3} - m_1 \\ k_{22}\frac{X_2}{X_3} + k_{23}\frac{1}{X_3} - m_2 \end{bmatrix}.$$

The residual vector in camera 2 is the vector from the projection of $\mathbf{X}$ to $\mathbf{m}$, so

$$\mathbf{e}_2(\mathbf{X}) = \begin{bmatrix} k_{11}'\frac{\mathbf{r}_1^T(\mathbf{X}-\mathbf{t})}{\mathbf{r}_3^T(\mathbf{X}-\mathbf{t})} + k_{12}'\frac{\mathbf{r}_2^T(\mathbf{X}-\mathbf{t})}{\mathbf{r}_3^T(\mathbf{X}-\mathbf{t})} + k_{13}'\frac{1}{\mathbf{r}_3^T(\mathbf{X}-\mathbf{t})} - m_1' \\ k_{22}'\frac{\mathbf{r}_2^T(\mathbf{X}-\mathbf{t})}{\mathbf{r}_3^T(\mathbf{X}-\mathbf{t})} + k_{23}'\frac{1}{\mathbf{r}_3^T(\mathbf{X}-\mathbf{t})} - m_2' \end{bmatrix}.$$

The *reprojection error* $E_r(\mathbf{X})$ for a point $\mathbf{X}$ in 3D is the sum of distances in each camera from the projections of the point to the measured locations, so

$$E_r(\mathbf{X}) = \mathbf{e}_1^T\mathbf{e}_1 + \mathbf{e}_2^T\mathbf{e}_2.$$

It is natural to obtain $\mathbf{X}$ by simply minimizing the reprojection error.

Procedure: 31.4 *Triangulating by minimizing reprojection error*

Start with a point $\mathbf{c}$ viewed in a calibrated camera and a corresponding point $\mathbf{c}'$ in a second calibrated camera. The rotation $\mathcal{R}$ and translation $\mathbf{t}$ from the first to the second camera are known. The first camera's intrinsic matrix is $\mathcal{K}$ and the second camera's is $\mathcal{K}'$. Compute the reprojection error $E_r(\mathbf{X})$ for a variable 3D point $\mathbf{X}$ and minimize

$$E_r(\mathbf{X})$$

as a function of $\mathbf{X}$. Use a quasi-newton method for minimization.

31.2.3 The Essential Matrix

The first camera has focal point at the origin, so $\mathbf{v}_1 = \mathbf{X}$ is the vector from $\mathbf{f}_1$ to $\mathbf{X}$. The vector from $\mathbf{f}_2$ to $\mathbf{X}$ is $\mathbf{v}_2 = \mathbf{X} - \mathbf{t}$ *in the first camera's coordinate system.* From Figure 31.3 $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{t}$ must be coplanar. This means that

$$[\mathbf{t} \times \mathbf{v}_1]^T \mathbf{v}_2 = 0$$

A convenient trick from linear algebra helps here. For a vector $\mathbf{a} = [a_1, a_2, a_3]^T$, write

$$[\mathbf{a}]_X = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}$$

and notice $\mathbf{a} \times \mathbf{b} = [\mathbf{a}]_X \mathbf{b}$. Further, you can write $\mathbf{v}_2$ in the second camera's coordinate system, yielding $\mathbf{v}_2' = \mathcal{R}^T \mathbf{v}_2$. This means that

$$\begin{aligned} [[\mathbf{t}]_X \mathbf{v}_1]^T \mathbf{v}_2 &= 0 \\ &= \mathbf{v}_1^T [\mathbf{t}]_X{}^T \mathbf{v}_2 \\ &= \mathbf{v}_1^T [\mathbf{t}]_X{}^T \mathcal{R} \mathbf{v}_2' \end{aligned}$$

Camera 1's image plane is at $z = 1$ in that camera's coodinate system, so the point $\mathbf{x}_1$ is at $(x_{1,1}, x_{2,1}, 1)^T$ in 3D and in camera 1's coordinate system. Alternatively, $(x_{1,1}, x_{2,1}, 1)^T$ is the location of $\mathbf{x}_1$ on camera 1's image plane in homogeneous coordinates. Further, $\mathbf{v}_1 \equiv \mathbf{x}_1$. Express both $\mathbf{x}_1$ and $\mathbf{x}_2'$ in homogeneous coordinates. Then

$$\begin{aligned} \mathbf{x}_1^T [\mathbf{t}]_X{}^T \mathcal{R} \mathbf{x}_2' &= \mathbf{v}_1^T [\mathbf{t}]_X{}^T \mathcal{R} \mathbf{v}_2' \\ &= 0 \end{aligned}$$

and so

$$\mathbf{x}_1^T \mathcal{E} \mathbf{x}_2' = 0$$

where $\mathcal{E}$ is known as the *essential matrix.*

Remember this: *Given two cameras related by a rotation $\mathcal{R}$ and a translation $\mathbf{t}$, the essential matrix is*

$$\mathcal{E} = [\mathbf{t}]_X{}^T \mathcal{R}.$$

A point $\mathbf{X}$ in 3D projects to $\mathbf{x}_1$ on camera 1's image plane in camera 1's coordinate system. The same point projects to $\mathbf{x}_2'$ on camera 2's image plane in camera 2's coordinate system. For any $\mathbf{X}$ not on the baseline,

$$\mathbf{x}_1^T \mathcal{E} \mathbf{x}_2' = 0$$

31.2.4 Properties of the Essential Matrix

The singular values of the essential matrix are the same as the singular values of $[\mathbf{t}]_X$ **exercises**. This means that one singular value must be zero, because

$$\mathbf{t}^T [\mathbf{t}]_X = \mathbf{0}^T.$$

Further, the remaining two singular values must be the same. For any vector $\mathbf{v}$ that is perpendicular to $\mathbf{t}$,

$$\| \mathbf{t} \times \mathbf{v} \| = \| \mathbf{t} \| \| \mathbf{v} \|.$$

In turn, for any such vector

$$\| [\mathbf{t}]_X \mathbf{v} \| = \| \mathbf{t} \| \| \mathbf{v} \|.$$

There is a two dimensional space of such vectors, so that there is a two dimensional space of vectors $\mathbf{v}$ so that

$$\frac{\| [\mathbf{t}]_X \mathbf{v} \|}{\| \mathbf{v} \|} = \| \mathbf{t} \|$$

and so the remaining two singular values must be the same **exercises**.

Remember this: *The essential matrix has one singular value zero, The two remaining singular values are equal.*

31.2.5 Estimating the Essential Matrix

The essential matrix can be obtained *up to scale* from the fundamental matrix if you know the intrinsic calibration of the two cameras. The calibration matrices are written $\mathcal{K}_1$ and $\mathcal{K}_2$, and must have full rank. Then you can estimate the essential matrix as

$$\mathcal{K}_1^{-T} \mathcal{F} \mathcal{K}_2^{-1}.$$

Since the fundamental matrix is meaningful only up to scale, this estimate is only up to scale as well. This means that, if you apply the constructions above for the fundamental matrix to the essential matrix, they yield epipolar lines or epipoles *in the world coordinate system of the relevant camera*. The essential matrix must have rank 2 because the fundamental matrix does, but it must also have two equal singular values, and so this estimate must be corrected. Because the estimate is meaningful only up to scale, you can choose these singular values to be one.

Procedure: 31.5 *Estimating the essential matrix*

Use procedure 31.3 to estimate the fundamental matrix, then use the camera intrinsics to compute

$$\mathcal{M} = \mathcal{K}_1^T \mathcal{F} \mathcal{K}_2.$$

From $\mathcal{M}$ compute the SVD to obtain $\mathcal{U}$, Σ and $\mathcal{V}$. Write $\Sigma_e = \text{diag}(1, 1, 0)$. Then the estimate of the essential matrix is

$$\hat{\mathcal{E}} = \mathcal{U} \Sigma_e \mathcal{V}^T.$$

Estimating the essential matrix like this is, in some sense, wasteful. It is known how to estimate an essential matrix using as few as five point correspondences [].

31.3 VISUAL ODOMETRY: EXPLOITING AN ESSENTIAL MATRIX

Remarkably, an essential matrix reveals the transformation between the two cameras up to scale. Recovering this information involves slightly complicated geometry. The fact that, given two images from cameras with known calibration, you can determine the relations between the cameras, is useful. For example, you can use it to estimate the motion of vehicles or drones from images.

31.3.1 Recovering Translation up to Scale

Assume you have two images obtained from two cameras whose calibration matrices you know. Further, you have the fundamental matrix up to scale, using the procedures of Section **??**. You can recover the essential matrix without difficulty, as in Procedure 31.5.

The essential matrix you estimate $\hat{\mathcal{E}}$ is

$$\hat{\mathcal{E}} = s[\mathbf{t}]_X \mathcal{R}$$

(for some unknown $s \neq 0$). Further $\mathbf{t}^T [\mathbf{t}]_X = \mathbf{0}$, so you can immediately recover the translation $\mathbf{t}$ *up to scale* by finding the unit vector $\mathbf{u}$ such that

$$\mathbf{u}^T \hat{\mathcal{E}} = \mathbf{0}^T.$$

This vector is occasionally referred to as the *left null vector* of the essential matrix. It is an estimate up to scale of the translation $\mathbf{t}$. Notice there are two unit left null

vectors because $-\mathbf{u}$ is also a unit vector and is also a left null vector. Notice there is also a unit vector $\mathbf{s}$ such that

$$\hat{\mathcal{E}}\mathbf{s} = \mathbf{0}.$$

This is the *right null vector* of the essential matrix; again, there are two.

31.3.2 Recovering the Rotation

Choose one of the left null vectors, and call it $\mathbf{u}$. Then

$$\hat{\mathcal{E}} = [\mathbf{u}]_X \mathcal{R}$$

for an unknown $\mathcal{R}$. Take the singular value decomposition of $[\mathbf{u}]_X$ to get

$$[\mathbf{u}]_X = \mathcal{U}_u \Sigma_e \mathcal{V}_u^T$$

where $\Sigma_e = \text{diag}(1, 1, 0)$ **exercises**. Take the singular value decomposition of $\hat{\mathcal{E}}$ to get

$$\hat{\mathcal{E}} = \mathcal{U}_e \Sigma_e \mathcal{V}_e^T.$$

Now

$$\hat{\mathcal{E}} = [\mathbf{u}]_X \mathcal{R} = \mathcal{U}_u \Sigma_e \mathcal{V}_u^T \mathcal{R} = \mathcal{U}_e \Sigma_e \mathcal{V}_e^T.$$

Conclude that

$$\mathcal{R} = \mathcal{V}_u \mathcal{V}_e^T.$$

However, there are important ambiguities, because Σ_e does not have full rank.

31.3.3 Ambiguities in the Odometry Estimate

Because $[\mathbf{u}]_X \mathbf{u} = \mathbf{0}$

$$\mathcal{V}_u^T \mathbf{u} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

(**exercises**) so you can write

$$\mathcal{V}_u = [\mathbf{a}_1 \mathbf{a}_2 \mathbf{u}].$$

Because $\hat{\mathcal{E}}\mathbf{s} = \mathbf{0}$

$$\mathcal{V}_e^T \mathbf{s} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

(**exercises**) so you can write

$$\mathcal{V}_e = [\mathbf{b}_1 \mathbf{b}_2 \mathbf{s}].$$

Now multiply out these two expressions to obtain

$$\mathcal{R} = \mathbf{a}_1 \mathbf{b}_1^T + \mathbf{a}_2 \mathbf{b}_2^T + \mathbf{u}\mathbf{s}^T$$

The sign ambiguity means there are two possible versions of $\mathbf{u}$ and two possible versions of $\mathbf{s}$. However, there are only two possible versions of $\mathbf{u}\mathbf{s}^T$ (two negatives are the same as two positives, **exercises**). Now construct the two matrices

$$\mathcal{W}_+ = \mathbf{a}_1\mathbf{b}_1^T + \mathbf{a}_2\mathbf{b}_2^T + \hat{\mathbf{u}}\hat{\mathbf{s}}^T$$

and

$$\mathcal{W}_- = \mathbf{a}_1\mathbf{b}_1^T + \mathbf{a}_2\mathbf{b}_2^T - \hat{\mathbf{u}}\hat{\mathbf{s}}^T.$$

and write

$$\hat{\mathcal{R}}_+ = \mathcal{W}_+(\det(\mathcal{W}_+))$$
$$\hat{\mathcal{R}}_- = \mathcal{W}_-(\det(\mathcal{W}_-)).$$

These are (a) true rotations, because their determinants are positive and (b) estimates of the rotation. Recall that there are two possible estimates of the translation, $\hat{\mathbf{u}}$ and $-\hat{\mathbf{u}}$. Then the essential matrix yields four possible *distinct* camera configurations. One of

$$(\hat{\mathbf{u}}, \hat{\mathcal{R}}_+), (-\hat{\mathbf{u}}, \hat{\mathcal{R}}_+), (\hat{\mathbf{u}}, \hat{\mathcal{R}}_-), (-\hat{\mathbf{u}}, \hat{\mathcal{R}}_-)$$

is the correct pair of (scaled translation, rotation).

Procedure: 31.6 *Estimating camera rotation and translation from an essential matrix*

Given $\hat{\mathcal{E}}$, an essential matrix, construct $\hat{\mathbf{u}}$, one of the two available unit left null vectors. This is an estimate of the translation up to scale. Take the singular value decomposition of $[\mathbf{u}]_X$ to get $[\mathbf{u}]_X = \mathcal{U}_u\Sigma_e\mathcal{V}_u^T$ where $\Sigma_e = \text{diag}(1, 1, 0)$. Take the singular value decomposition of $\hat{\mathcal{E}}$ to get $\hat{\mathcal{E}} = \mathcal{U}_e\Sigma_e\mathcal{V}_e^T$. Now $\mathcal{V}_u = [\mathbf{a}_1\mathbf{a}_2\mathbf{u}]$ and $\mathcal{V}_e = [\mathbf{b}_1\mathbf{b}_2\mathbf{s}]$. Construct the two matrices $\mathcal{W}_+ = \mathbf{a}_1\mathbf{b}_1^T + \mathbf{a}_2\mathbf{b}_2^T + \hat{\mathbf{u}}\hat{\mathbf{s}}^T$ and $\mathcal{W}_- = \mathbf{a}_1\mathbf{b}_1^T + \mathbf{a}_2\mathbf{b}_2^T - \hat{\mathbf{u}}\hat{\mathbf{s}}^T$. Write $\hat{\mathcal{R}}_+ = \mathcal{W}_+(\det(\mathcal{W}_+))$ and $\hat{\mathcal{R}}_- = \mathcal{W}_-(\det(\mathcal{W}_-))$. The essential matrix yields four possible camera configurations:

$$(\hat{\mathbf{u}}, \hat{\mathcal{R}}_+), (-\hat{\mathbf{u}}_-, \hat{\mathcal{R}}_+), (\hat{\mathbf{u}}, \hat{\mathcal{R}}_-), (-\hat{\mathbf{u}}_-, \hat{\mathcal{R}}_-).$$

31.3.4 Disambiguating Reconstructions

The relations between $\hat{\mathcal{R}}_+$ and $\hat{\mathcal{R}}_-$ are revealing. Compute

$$\hat{\mathcal{R}}_-\hat{\mathcal{R}}_+^T = -\mathbf{a}_1\mathbf{a}_1^T - \mathbf{a}_2\mathbf{a}_2^T + \hat{\mathbf{u}}\hat{\mathbf{u}}^T$$

and this maps $\mathbf{u}$ to $\mathbf{u}$, $\mathbf{a}_1$ to $-\mathbf{a}_1$ and $\mathbf{a}_2$ to $-\mathbf{a}_2$ – it is a rotation by 180^0 around the axis $\hat{\mathbf{u}}$. Now think about triangulating using each of the four reconstructions. Figure 31.4 visualizes the effects. Choose camera 1, and use the procedure to come up with four possible configurations for camera 2 (2(a)...2(d)). Each of the reconstructions of camera 2 *has the same essential matrix* with camera 1, so that

if a point in one reconstruction corresponds to a point in camera 1, so does that point in each reconstruction. In turn, this means you can triangulate the point in each reconstruction with the point in camera 1. This leads to four distinct points in space, shown in the figure. In this figure, only one of the four triangulated points lies in front of both camera 1 and a reconstruction. This is the general case. You can choose the correct reconstruction by this property, which you can test by looking at the sign of the Z-coordinate *in each camera's frame* **exercises**.

Procedure: 31.7 *Disambiguating odometry solutions*

Construct the four solutions of Procedure 31.6, yielding four distinct cameras. Pair each of these cameras with the unit camera (which is camera 1) and for each of these four pairs, triangulate a set of points. Ideally, in one pair, the points will all be in front of both cameras. In practice, error in localizing the points might lead some to be behind one camera, so choose the reconstruction where the largest fraction of reconstructions is in front of both cameras.

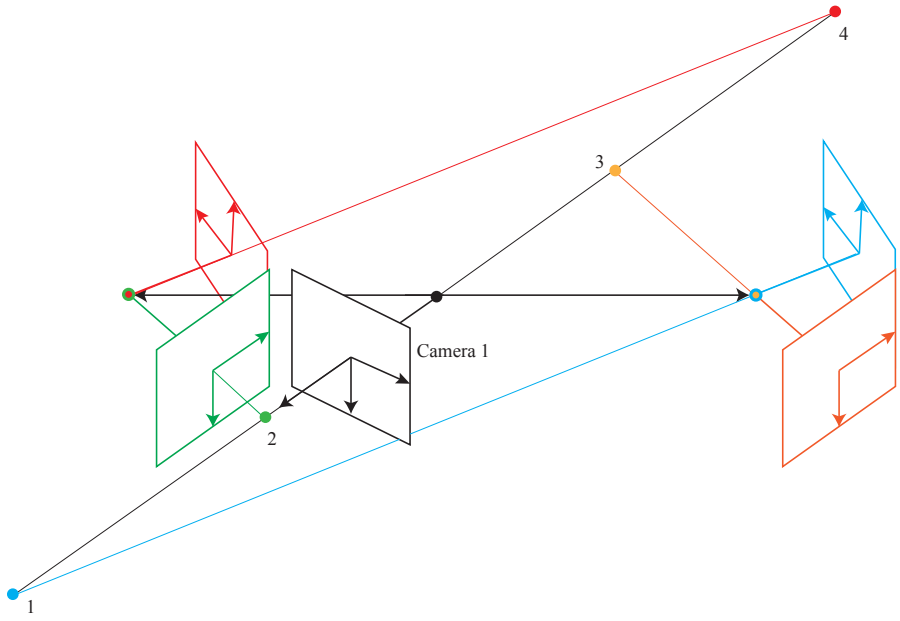


FIGURE 31.4: *There are four distinct possible reconstructions for the transformation from camera 1 to camera 2 implied by a known essential matrix. In this figure, camera 1 is at the canonical location; the black point at the center of the figure is camera 1's focal point; and the red, green, blue and orange cameras are the four reconstructions. Notice how red and green are related by a* 180^0 *rotation about the translation vector, as are blue and orange. Red and green share a focal point (red with green outline) and blue and orange share a focal point (orange with blue outline). The red-green pair are associated with one sign for the translation estimate (the black arrow), and the blue-orange pair are associated with the other sign. Each reconstruction has a triangulation of a point associated with it. I have marked the point with the same color as the camera 2 reconstruction. Notice how only one triangulation – in this case, the green point – produces a point that lies in front of both camera 1 and the reconstruction. This is the general case. The green camera must be the correct reconstruction.*

31.4 YOU SHOULD

31.4.1 remember these facts:

The fundamental matrix 573
The fundamental matrix has rank 2 575
The fundamental matrix is only meaningful up to scale 575
The essential matrix 580
Singular values of the essential matrix 580

31.4.2 remember these procedures:

Obtaining an epipolar line from a fundamental matrix 574
Obtaining epipoles from a fundamental matrix 574
Estimating the fundamental matrix 576
Triangulating by minimizing reprojection error 579
Estimating the essential matrix 581
Estimating camera rotation and translation from an essential matrix 583
Disambiguating odometry solutions 584

31.4.3 be able to:

- Estimate movements approximately from epipoles.
- Estimate a fundamental matrix.
- Extract epipoles from a fundamental matrix.
- Extract epipolar lines from a fundamental matrix.
- Triangulate a point in 3D from two images in calibrated cameras.
- Estimate an essential matrix.
- Extract odometry information from an essential matrix estimate.

EXERCISES

QUICK CHECKS

31.1. A point in 3D has homogeneous coordinates $\mathbf{X}$. This point lies on the plane $\mathbf{A}^T\mathbf{X} = 0$, but nothing else is known about the point. Show that there is a 4×3 matrix $\mathcal{M}$ and 3×1 vector of homogeneous coordinates such that $\mathbf{X} = \mathcal{M}\mathbf{u}$ (**hint:** the columns $\mathbf{m}_i$ of the matrix have the property $\mathbf{A}^T\mathbf{m}_i = 0$).

31.2. Section 31.1.3 has:"In fact, $\mathcal{F}$ must have rank 2. The rank can't be three, because there are epipoles." Explain.

31.3. Section 31.1.3 has: "The fundamental matrix is a map from points in one camera to lines in the other." Explain.

31.4. Section 31.1.3 has: "The fundamental matrix is a map from points in one camera to lines in the other. If the rank were 0, the fundamental matrix would map any point in one camera to a zero vector (this happens if the camera is not translated, a rather special case, Section **??**)." Explain.

31.5. Section 31.2.4 has: "The singular values of the essential matrix are the same as the singular values of $[\mathbf{t}]_X$." Explain.

31.6. For $\mathbf{u}$ a unit vector, show that the singular values of $[\mathbf{u}]_X$ are 1, 1, and 0.

31.7. Section 31.3.3 has: "Because $[\mathbf{u}]_X\mathbf{u} = \mathbf{0}$

$$\mathcal{V}_u^T\mathbf{u} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}."$$

Explain.

31.8. Section 31.3.3 has: "Because $\hat{\mathcal{E}}\mathbf{s} = \mathbf{0}$

$$\mathcal{V}_e^T\mathbf{s} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}."$$

Explain.

31.9. Section 31.3.3 has: "The sign ambiguity means there are two possible versions of $\mathbf{u}$ and two possible versions of $\mathbf{s}$. However, there are only two possible versions of $\mathbf{us}^T$ (two negatives are the same as two positives)." Explain.

31.10. Section 31.3.4 has: "You can choose the correct reconstruction by this property, which you can test by looking at the sign of the Z-coordinate *in each camera's frame*." Explain.

LONGER PROBLEMS

31.11. This exercise explores the relationship between determinants and linear spaces.

(a) You have three points on the plane, given in homogeneous coordinates by

$$\mathbf{X}_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \text{ and } \mathbf{X}_2 = \begin{bmatrix} a \\ 0 \\ 1 \end{bmatrix} \text{ and } \mathbf{X}_3 = \begin{bmatrix} b \\ c \\ 1 \end{bmatrix}$$

Show that the area of the triangle they subtend is proportional to

$$\text{determinant}([\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3])$$

(b) You have three points on the plane. The i'th point has homogeneous coordinates $\mathbf{X}_i$. Show that

$$\text{determinant}([\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3]) = 0$$

if and only if the three points are collinear. **Hint:** Use the results of the previous exercise.

(c) You have four points in 3D. The i'th point has homogeneous coordinates $\mathbf{X}_i$. Show that

$$\text{determinant}([\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4]) = 0$$

if and only if the four points are coplanar. **Hint:** Do the previous two subexercises first.

31.12. Section 31.2.4 has: " Further, the remaining two singular values must be the same. For any vector $\mathbf{v}$ that is perpendicular to $\mathbf{t}$,

$$\|\mathbf{t} \times \mathbf{v}\| = \|\mathbf{t}\|\|\mathbf{v}\|.$$

In turn, for any such vector

$$\|[\mathbf{t}]_X \mathbf{v}\| = \|\mathbf{t}\|\|\mathbf{v}\|.$$

There is a two dimensional space of such vectors, so that there is a two dimensional space of vectors $\mathbf{v}$ so that

$$\frac{\|[\mathbf{t}]_X \mathbf{v}\|}{\|\mathbf{v}\|} = \|\mathbf{t}\|$$

and so the remaining two singular values must be the same." Explain.

PROGRAMMING EXERCISES

31.13. The tanks and temples dataset is a standard dataset of multiple views of scenes, available at `https://www.tanksandtemples.org/download/`. Obtain the frames in the "Family" sequence, and obtain a copy of the Superpoint pretrained network, released at `https://github.com/magicleap/SuperPointPretrainedNetwork`, and a small collection of images.

(a) Use the Superpoint network to obtain a set of putative correspondences between frames 0001.jpg and 0002.jpg of the sequence.

(b) Use Ransac and the eight point algorithm to estimate the fundamental matrix between 0001.jpg and 0002.jpg of the sequence. Do not scale or otherwise adjust the image coordinates.

(c) Use Ransac and the eight point algorithm to estimate the fundamental matrix between 0001.jpg and 0002.jpg of the sequence, but now scale the image coordinates to range from 0-1 along the longest axis of the image.

(d) Which fundamental matrix estimate is better? Why do you claim that it is better?

(e) Now use your best procedure to estimate the fundamental matrix between 0001.jpg and 0017.jpg. Do you think this estimate is reliable? Why?